

¿Cómo citar el artículo?

Bedoya Herrera, O. M., López Trujillo, M. y Marulanda Echeverry, C. E. (septiembre-diciembre, 2019). Modelo predictivo para la identificación de factores socioculturales asociados al tiempo de búsqueda del primer empleo en egresados universitarios. *Revista Virtual Universidad Católica del Norte*, (58), 3-18. doi: <https://doi.org/10.35575/rvucn.n58a6>

| Modelo predictivo para la identificación de factores socioculturales asociados al tiempo de búsqueda del primer empleo en egresados universitarios ¹

Predictive model for the identification of factors associated with the search time of the first job in university graduates

Oscar Mauricio Bedoya Herrera

Magister en Ingeniería Computacional
Universidad de Caldas, Manizales
Colombia
oscar.bedoya@ucaldas.edu.co
Orcid: <http://orcid.org/0000-0002-6757-3327>

Marcelo López Trujillo

Doctor en Ingeniería Informática
Universidad de Caldas, Manizales
Colombia
mlopez@ucaldas.edu.co
Orcid: <http://orcid.org/0000-0003-0668-1292>

Carlos Eduardo Marulanda Echeverry

Doctorado en Ingeniería: Industria y organizaciones
Universidad de Nacional, Manizales
Colombia
cemarulandae@unal.edu.co
Orcid: <http://orcid.org/0000-0002-6287-9386>

Recibido: 02 de mayo de 2018

Evaluado: 10 de junio de 2019

Aprobado: 08 de julio de 2019

Tipo de artículo: Investigación científica y tecnológica.

¹ Proyecto de investigación titulado: Explotación de datos masivos en la bodega de datos de egresados de la Universidad de Caldas y su integración con procesos de gestión del conocimiento e inteligencia de negocios. Grupo GITIR, Departamento de Sistemas e Informática, Universidad de Caldas, Calle 65 No. 26-10 Manizales, Colombia.

| Resumen

Algunos programas de educación superior en Colombia desconocen si las variables socioculturales de sus recién egresados son influyentes en el tiempo que tardan en conseguir su primer empleo. Este artículo muestra los resultados del uso de técnicas automáticas de extracción de datos, para el análisis predictivo del tiempo que un graduado tarda en conseguir su primer empleo, orientadas a egresados de Ingeniería de Sistemas de la Universidad de Caldas, en el último lustro. Para su desarrollo se emplearon rutinas de análisis de datos, utilizando el algoritmo de longitud de descripción mínima, estableciendo que los factores representativos, que permiten pronosticar el tiempo de búsqueda satisfactoria del primer empleo, son el nivel de educación de la madre, el rendimiento académico y el factor de ingreso. El estudio contrasta el valor objetivo original con el valor de predicción para 135 registros, donde se concluye que los graduados que tienen un rendimiento de 3, entre 1 y 5, siendo este último el rendimiento ideal, con un nivel de educación de su madre correspondiente a "Primaria Completa", y cuyo factor de ingreso es "Sus habilidades y destrezas", se proyectan para conseguir su primer empleo en "Menos de 3 meses", con un 63% de certeza.

Palabras clave: Análisis predictivo, Construcción del modelo, Egresados, Minería de datos, Técnicas de predicción.

| Abstract

Some higher education programs in Colombia do not know if the sociocultural variables of their recent graduates are influential in the time it takes to get their first job. This article shows the results in the use of automatic techniques of data extraction for the predictive analysis of the time that a graduate takes to get his first job oriented to graduates of Systems Engineering of the University of Caldas in last five years. For its development, data analysis routines were used using the minimal description length algorithm, establishing that the representative factors that allow predicting the satisfactory search time of the first job are the level of education of the mother, the academic performance and the factor of income. The study contrasts the original objective value with the prediction value for 135 records where it is concluded that the graduates have a performance of 3 between 1 and 5, the latter being the ideal performance, with a level of education of their mother corresponding to "Primary Complete" whose income factor is "Your skills and abilities", is projected to get your first job in "Less than 3 months", with 63% certainty.

Keywords: Prediction techniques, Construction of the model, Graduates, Data mining, Predictive analysis

| Introducción

Las instituciones de educación superior cumplen un rol determinante dentro de la sociedad: educan jóvenes profesionales con una formación teórica, técnica, de competencias y habilidades frente a una ciencia específica. Buron y Saussen (2017) mencionan que las universidades en sí contemplan una propuesta de formar jóvenes para el mercado de trabajo, a través del conocimiento adquirido y lapidado dentro de ella misma. No obstante, el desempleo, la poca sostenibilidad del recién egresado y la dificultad para conseguir una vacante, han llevado a las universidades del país a investigar las causas que subyacen esta problemática.

La minería de datos surge, entonces, como una solución para descubrir información y conocimiento, y las relaciones existentes entre las diversas variables que se pueden estudiar, desde los métodos estadísticos tradicionales y los procesos automáticos de extracción de datos. En el caso de la educación universitaria, las técnicas de minería de datos procuran indagar, explorar y analizar la información de los registros de bases de datos de diversas fuentes, que pueden ir desde los docentes, estudiantes y administrativos, o como el caso de esta investigación, minar la información de los recién egresados, relacionada con en el tiempo ocasionado para su primer ejercicio profesional, y que constituye parte vital en el desarrollo integral del individuo, además de ser una de las principales causas de retraso económico que vive el país.

La Universidad de Caldas, en Manizales-Colombia, no es ajena a esta situación; en este sentido, se presentan los resultados de la investigación, con el objeto de aportar conocimiento que contribuya a establecer una visión, en procura de predecir los factores socio culturales que influyen en el índice del tiempo de inserción laboral de un graduado, teniendo en cuenta el conjunto de datos de los egresados del último lustro del programa de Ingeniería de Sistemas. En esta investigación se pretende generar conocimiento, orientado a determinar qué tanto influyen las variables socioculturales, de rendimiento académico y familiares en el tiempo que un egresado tarda en conseguir su primer empleo.

La minería de datos es un factor determinante en el proceso conocido como generación de conocimiento. En efecto, Natek & Zwilling (2014) explican que las técnicas de minería de datos, usadas en el proceso de gestión del conocimiento, pueden ser utilizadas para extraer y descubrir información apreciable y explicativa de una gran cantidad de datos.

Al respecto, Shaw, Subramaniam, Tan & Welge (2001) consideran la minería de datos como el proceso de exploración de análisis de datos para revelar información oculta y potencialmente valiosa para la organización. Los autores señalan tres ámbitos principales de aplicación de las herramientas de minería de datos: generación de perfiles de actores involucrados, análisis de desviaciones y análisis de tendencias.

Por otra parte, el proceso de gestión de conocimiento, enmarcado en procesos de minería de datos, viene determinado por la extracción de patrones desconocidos que son descubiertos por el análisis automático o semi-automático de datos. Estos patrones pueden ser grupos de registros (análisis clúster), registros inusuales (detección de anomalías) o también formar dependencias entre datos (asociación reglas). Al final, los patrones pueden ser vistos como un resumen de los datos de entrada, y se pueden utilizar para su posterior análisis y beneficio.

En este sentido, Hand, Mannila & Smyth (2001) explican que la minería de datos proporciona una alternativa de solución para el análisis de los fenómenos no explícitos en bases de datos; estos autores caracterizan la minería de datos como un ejercicio interdisciplinario, que involucra estadística, tecnologías de bases de datos, inteligencia artificial, reconocimiento de patrones y visualización de datos.

Como muestra, existen múltiples espacios de aplicación para los procesos de minería de datos que generan conocimiento valioso; algunos ejemplos importantes, señalados en estudios de investigación, son: en el área de la educación, Martelo, Acevedo y Martelo (2018) exponen una investigación que tiene por objetivo determinar los factores que inciden en la deserción de estudiantes del programa de Ingeniería de Sistemas de la Universidad de Cartagena, mediante análisis multivariado. Así mismo, M. Miranda & Guzmán (2017) muestran un estudio para determinar las variables que producen el abandono de la entidad educativa universitaria, por parte de un estudiante, usando técnicas de minería de datos, donde se concluye y determina que las variables que mejor advierten la deserción de un estudiante son: las razones socioeconómicas y el puntaje de ingreso a la universidad.

En esta misma línea, vale decir que Bedoya, M. López y Marulanda (2016) comparan distintos algoritmos de minería de datos y evidencian los factores resultantes con técnicas de clasificación de minería de datos, para determinar la predicción de la utilidad de los conocimientos en graduandos de la Universidad de Caldas. También, A. Miranda y Tirado (2012) emplean técnicas de minería de datos para analizar la actividad mediada de una comunidad virtual de aprendizaje. De igual modo, Aponte, Hoyos y Monsalve (2012) estudian diferentes técnicas, herramientas y enfoques de la minería de datos, en el análisis del uso dado por estudiantes y docentes a las plataformas virtuales de aprendizaje, que sirven como apoyo al desarrollo de procesos de enseñanza.

De otro lado, en sectores de tecnologías de inteligencia artificial, Solarte y Millán (2014) describen una propuesta para extender el proceso de recuperación de información, usando tecnologías de la web semántica con minería de datos. También, en el área de la medicina, Sánchez et al. (2018) ilustran el valor agregado que aportan los modelos de minería de datos en la toma de decisiones, para determinar la malignidad de una masa detectada en el seno de un paciente.

Con lo anterior, se podría afirmar que en la minería de datos, en cualquier disciplina, tanto las aplicaciones como las herramientas operan de forma conjunta, con el propósito de generar conocimiento que aporte seguridad en la toma de decisiones; no obstante, como lo anotan Marulanda, López y Mejía (2017) las técnicas de minería de datos implican, en sus primeras fases, limpieza, ajuste y transformación, de acuerdo a las necesidades del estudio; los autores aseveran que dichos procesos garantizan los posibles errores del banco de datos, los cuales podrían incurrir en desaciertos para la coincidencia de datos y la futura gestión del conocimiento.

Por otra parte, las funcionalidades de minería de datos pueden ser divididas en varios grupos, como son: la caracterización y discriminación, la extracción de patrones y reglas de asociación, la clasificación y predicción, y el análisis de conglomerados. Los métodos de clasificación predicen una clase discreta de datos, mientras que el método de predicción sus funciones.

La clasificación también se conoce como aprendizaje supervisado, ya que requiere que el número de clases se defina de antemano y necesita un conjunto de datos de entrenamiento con etiquetas predefinidas. Es un proceso de dos pasos: el primer paso es construir clasificadores basados en datos, y el segundo paso utiliza los clasificadores establecidos para pronosticar etiquetas de clase de datos desconocidos; en este orden de ideas, Han, Kamber & Pei (2012) establecen que la minería de datos se fundamenta en el conocimiento de los patrones ocultos y las asociaciones, en procura de la construcción de criterios de análisis, en la clasificación y la predicción, mostrando los resultados y los beneficios con herramientas de visualización.

Complementariamente, Bedoya, D. López y Marulanda (2018) manifiestan que las categorías básicas, usadas en las técnicas de minería de datos, se pueden reunir en la clasificación, la asociación, la secuenciación y la clusterización. El método de clasificación de individuos en grupos permite descubrir reglas y, por ende, asemejar cuáles son las variables que mejor puntualicen la pertenencia un determinado grupo. La asociación permite establecer las conexiones entre uno o más eventos, evaluando correlaciones. Conviene además los autores que en el análisis clúster se permite clasificar un conglomerado en un número categórico de grupos, con base a similitudes y disimilitudes de perfiles existentes entre los diferentes componentes de dicha conglomerado. Lo anterior se ratifica con lo establecido por Gómez y Suárez (2010), quienes advierten que la segmentación y agrupamiento permiten llevar a cabo la identificación de grupos con características similares, lo que conlleva a describir patrones en cada segmento, para su efectivo uso.

Otro aspecto importante consiste en el proceso de minería de datos; según Timaran (2010) la minería de datos es un desarrollo automático en el que se combinan descubrimiento y análisis; dichas composiciones se convierten en un proceso iterativo e interactivo. Plantea el autor que la salida de alguna de las fases puede hacer retroalimentación en pasos anteriores, lo que conllevaría a convertir el proceso en iterativo. Asimismo, se convertiría en interactivo porque es el experto quien interviene en la toma de decisiones. Continúa el autor especificando que las etapas de la minería de datos se especializarían en:

Etapa de Interpretación y evaluación de resultados: En esta etapa se interpretan los patrones descubiertos y posiblemente se retorna a los anteriores pasos o etapas para posteriores iteraciones. Esta etapa puede incluir la visualización de los patrones extraídos (...) en términos que sean entendibles para el usuario. (Timaran, 2010, p. 127).

Ahora bien, considerando la minería de datos, orientada a generar conocimiento para la educación universitaria, Chalaris, Gritzalis, Maragoudakis, Sgouropoulou & Tsolakidis (2014) explican que hay una gran cantidad de técnicas de minería de datos que se pueden aplicar a los datos educativos, y cada uno de ellos puede proporcionar resultados útiles y resultados que ayudan a abordar muchos problemas en este ámbito. Las tareas de minería, como el agrupamiento, pueden revelar características comprensivas de los estudiantes, mientras que la predicción (clasificación y regresión) y la minería de relaciones (asociación, correlación, minería secuencial) pueden ayudar a la universidad en la formulación de políticas e iniciativas para disminuir la tasa de abandono escolar o aumentar la retención estudiantil, éxito y logro de resultados de aprendizaje. Podría ayudar a proporcionar una educación más personalizada, mayor eficiencia del sistema educativo y reducir el costo de los procesos educativos en cualquier país.

En suma, Medeiros y Galvão (2016) plantean un estudio sobre la asociación entre educación universitaria y riqueza; ellos examinan cómo la educación aumenta las posibilidades de un individuo de pertenecer al 1% más rico de la distribución de los ingresos, y los efectos de la educación universitaria general en la población.

En conclusión, los estudios de minería de datos en establecimientos reconocidos de enseñanza universitaria buscan establecer criterios, a través de la generación de conocimiento, que permitan anticipar dificultades en procura del mejoramiento de una región. Es así, como Rossetto y Gonçalves (2015) corroboran lo anterior, estableciendo una investigación donde se analiza el acceso a la enseñanza superior en Brasil y su impacto a partir de criterios de justicia distributiva, más específicamente, del enfoque de igualdad y de oportunidades, con patrones de datos, en cuidado de establecer rutas que permitan el mejoramiento de las clases menos favorecidas de su país.

Población muestral

Los datos analizados constan de una población de 4656 registros, categorizados en 11 campos, correspondientes a los alumnos egresados del programa de Ingeniería de Sistemas de la Universidad de Caldas, entre en los años 2012 y 2017. Esta información hace parte del registro de encuestas digitales, al momento que el estudiante se convierte en egresado. Adicionalmente, dichas encuestas fueron complementadas con el levantamiento de información, usando entrevistas telefónicas y visitas de campo, referente al tiempo que el egresado tardó en conseguir su primer empleo.

Comprender lo datos (etapa de selección)

De la generalidad de variables contenidas en las encuestas digitales se consideraron, para el análisis, aquellas que presentaban mayor particularidad en la validez de los datos; las mismas se muestran en la tabla 1.

Tabla 1
Atributos del repositorio de datos

Atributo	Descripción/Valores Categóricos
Id	Identificador único del registro
Estadocivil	Soltero, casado o en unión libre
Genero	Masculino o Femenino
Rangoedad	Menor de 18 años, entre 18 y 22 años, entre 23 y 25 años, mayor de 26 años.
Nivelpadre	Nunca estudió, primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, educación técnica, educación universitaria incompleta, educación universitaria completa, educación de postgrado
Nivelmadre	Ídem al nivel padre
Tiempoempleo	Tiempo que el egresado usó para para conseguir su primer empleo: menos de tres meses, entre tres y seis meses, más de seis y hasta 1 año o más de 1 año
Etnia	Especifica la raza originaria del egresado: mestizo, indio o afrocolombiano
Utileshabi	Determina si las habilidades adquiridas en el pregrado son valederas en su actividad profesional, desde: muy útiles, útiles, poco útiles o nada útiles
Performance	Indicador que establece el rendimiento del egresado a través de su historia como estudiante de pregrado. Valores Rendimiento Regular o Bajo, Rendimiento Aceptable, Rendimiento Bueno, Rendimiento Sobresaliente

Nota: elaboración propia.

En cuanto al rendimiento académico (PERFORMANCE), este proviene de una categorización numérica resultante del historial de notas del egresado, donde se mide el comportamiento, progreso y rendimiento, a lo largo del trasegar como estudiante de pregrado.

Etapas de procesamiento y transformación

Se incluye el paquete libre DBMS_PREDICTIVE_ANALYTICS, de Oracle Data Mining, cuya característica fundamental es la extracción automática de datos. Esta herramienta automatiza las siguientes etapas del proceso: preparar los datos para la minería y creación de modelos, utilizando los datos preparados y la calificación de nuevos datos. Su objetivo es el de encontrar patrones en la bodega de datos (población muestral), y así establecer correlaciones significativas que puedan utilizarse para predecir eventos futuros.

La función predictiva analiza los datos, encuentra patrones, y pronostica los resultados con un nivel de confianza asociado. La confianza es trascendental, ya que pondera la certeza de las predicciones. El rango de valores es de 0 a 1. Cuanto mayor sea el valor de confianza, más segura será la predicción.

El proceso se compone de las funciones: (EXPLAIN y PREDICT). EXPLAIN crea un modelo de importancia de los atributos, y analiza el conjunto de datos para determinar el valor explicativo de cada atributo. Cuanto mayor sea el valor explicativo para un atributo, más fuerte es la relación entre el mismo y el resultado. Seguidamente, el proceso retorna una lista de atributos clasificados en orden relativo de su impacto en la predicción. Esta información se deriva de los datos del modelo, en beneficio del patrón de atributo de importancia.

El algoritmo longitud mínima de la descripción ayuda a identificar los atributos con mayor influencia sobre el atributo objetivo. Para cada atributo a considerar de importancia, el resultado es un valor que determina una correlación explicativa de cada particularidad.

De otro lado, el proceso de predicción (PREDICT) se centra en el clasificador basado en el algoritmo Naive Bayes. Este algoritmo opera con un valor de exactitud o confianza predictiva. Cuantitativamente (y ésta es la gran aportación de los métodos bayesianos), da una medida probabilística de la importancia de esas variables en el problema (y por tanto una probabilidad explícita de las hipótesis que se formulan). Esta es quizá una de las diferencias fundamentales que ofrecen las redes bayesianas, con respecto a otros métodos.

Es decir, se determina la importancia relativa de los atributos en la predicción de un valor objetivo, conforme a una medida probabilística de importancia de las variables en la correlación, a modo que:

$$\text{Prob (B dado A)} = \text{Prob (A Y B)} / \text{Prob (A)}$$

El proceso determina automáticamente los tipos de datos, no crea objetos intermedios en el set de datos y produce una predicción para cada registro.

| Resultados

Seguidamente se presentan los resultados de las categorías y variables evaluadas, desde el ciclo de vida del modelo EXPLAIN, que hace posible predecir y valorar el tiempo del primer empleo. Los resultados del proceso EXPLAIN para el atributo (TIEMPOEMPLO) (tardanza en conseguir el primer empleo como egresado) son mostrados en la tabla 2.

Tabla 2

Modelo de importancia correlacional atributo Tiempo primer empleo

Attribute name	Explanatory value	Rank
Niveducmadre	0,6733	1
Performance	0,5078	2
Factoringreso	0,3053	3
Niveducpadre	0,0844	4
RangoEdad	0,0753	5
Etnia	0,0297	6
Genero	0,0183	7
Estadocivil	0,0316	8
Utileshabi	0,0131	9

Nota: elaboración propia.

Los resultados exponen que (NIVEDUCMADRE) es el mejor predictor para el atributo (TIEMPOEMPLO). El éxito en el tiempo del primer empleo (TIEMPOEMPLO) está estrechamente demarcado por el atributo correspondiente al nivel de educación de la madre (NIVEDUCMADRE), seguido por el rendimiento académico como estudiante (PERFORMANCE). La significancia del valor en el ranking originado por el (EXPLANATORY_VALUE) puede determinar valores cercanos de predicción y comportamientos según las correlaciones identificadas. Como se muestra en la figura 1, los valores más cercanos a 1 denotan la importancia correlacional de los mejores predictores para el atributo mencionado.

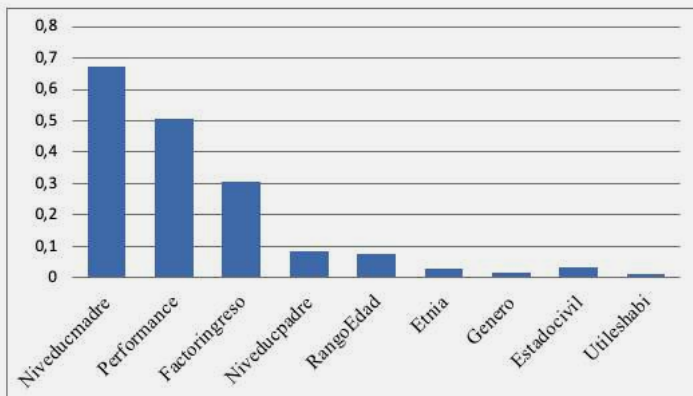


Figura 1. Importancia correlacional de los atributos frente al tiempo del primer empleo proceso Explain. Fuente: elaboración propia.

Adicionalmente, el estudio muestra los resultados EXPLAIN para el atributo (FACTORINGRESO) y (PERFORMANCE), como se muestra en las tablas 3 y 4, respectivamente.

Tabla 3
Modelo de importancia correlacional atributo Factor de ingreso

Attribute_name	Explanatory_value	Rank
Performance	0,15974355	1
Estadocivil	0,03992528	2
Niveducmadre	0,02513077	3
Niveducpadre	0,02162979	4
Genero	0,01188783	5
Tiempoempleo	0,00908048	6
Rangoedad	0,00790076	7
Utileshabi	0,00594687	8
Etnia	0,00594687	9

Nota: Elaboración propia.

Los resultados muestran que el atributo (PERFORMANCE) es el mejor predictor de la columna (FACTOR INGRESO); sin embargo, su significancia no corresponde a un valor predecible futuro.

En la tabla 4 se presentan los resultados para el proceso EXPLAIN, correspondiente al atributo (PERFORMANCE).

Tabla 4
Modelo de importancia correlacional atributo Performance

Attribute_name	Explanatory_value	Rank
Factoringreso	0,5605	1
Niveducmadre	0,4186	2
Tiempoempleo	0,1555	3
Niveducpadre	0,0766	4
Genero	0,0458	5
Rangoedad	0,0434	6
Estadocivil	0,0316	7
Utileshabi	0,0162	8
Etnia	0,0157	9

Nota: elaboración propia.

Los resultados muestran que (FACTORINGRESO) y (NIVEDUMADRE) son los mejores predictores de (PERFORMANCE).

Los tres campos que tienen un EXPLANATORY_VALUE cercanos a uno, son los atributos que se pueden utilizar para predecir el rendimiento (PERFORMANCE) que cada estudiante ha obtenido en su historia estudiantil. El factor de ingreso al programa de pregrado (FACTORINGRESO), registrado por el estudiante antes de graduarse, está más estrechamente alineado con el rendimiento, seguido de cerca por el nivel de educación de la madre (NIVEDUCMADRE).

El procedimiento PREDICT es el más útil en esta metodología; es usado para predecir los valores de una columna específica. La entrada consiste en unos parámetros y una columna de destino, la columna que contiene los valores objetivos de predecir. Los datos de entrada deben contener algunos casos en los que se conoce el valor objetivo (es decir, no es nulo). Los casos en los que se conocen los valores de destino se utilizan para entrenar y probar el modelo. PREDICT retornará un valor previsto para cada caso, incluyendo aquellos en los que se conoce el valor.

El objetivo será producir predicciones para objetivos desconocidos. Los casos donde el objetivo es desconocido, es decir, cuando el valor objetivo es nulo, no se tendrán en cuenta durante el entrenamiento del modelo.

Esta demostración de análisis predictivo partirá del procedimiento EXPLAIN, donde se determina que el atributo (TIEMPOEMPLEO) se correlaciona, en primera medida, con el atributo (NIVEDUCMADRE), seguido por el atributo (PERFORMANCE).

El siguiente proceso se enfocará en determinar el grado en que influye el atributo (NIVEDUCMADRE), en relación con la categorización del atributo (TIEMPOEMPLEO), utilizando un clasificador basado en el algoritmo Naïve Bayes.

La predicción de resultados se realiza para aquellos estudiantes cuya categorización del nivel de educación de la madre corresponde a "Secundaria completa"; finalmente, el proceso obtuvo una salida donde la tasa de valoración (ACCURACY) es de 74,02%, como se muestra en la tabla 5; valor muy representativo para la predicción.

El proceso de predicción PREDICT muestra que 288 egresados, cuyo nivel de educación de la madre es "Secundaria completa", según las correlaciones representativas socioeducativas de la bodega de datos, predicirán un tiempo de consecución del primer empleo de entre 7 meses y un año; 976 de ellos, "entre 3 meses y 6 meses"; y 3392, "Menos de 3 meses".

En efecto, los resultados del procedimiento PREDICT, para una medida de 250 registros, muestran que la probabilidad de predicción oscila entre un 50% y un 63%, valores determinantes para contrastar el modelo predictivo. Por tanto, para los egresados que tienen un rendimiento de 3, entre 1 y 5, siendo este último el rendimiento ideal, con un nivel de educación de su madre correspondiente a "Primaria Completa", y un nivel de educación de su padre de "Educación Técnica", que considera "Muy Útiles" los conocimientos adquiridos en la universidad, y cuyo factor de ingreso fue "Sus habilidades y destrezas", se proyecta en el tiempo para conseguir su primer empleo en "Menos de 3 meses", con un 63 por ciento (0,6343) de certeza.

| Discusión

Como se evidencia en el proceso, se encontró una relación significativa entre el nivel de educación de la madre y el tiempo del primer empleo; en donde los graduados que tardan más tiempo en conseguir empleo son aquellos cuyos padres no estudiaron o no terminaron la primaria. Cuando el nivel educativo de la madre es más alto, el tiempo que tarda el egresado en conseguir su primer trabajo disminuye. Sin embargo, el descubrimiento más importante fue la confianza de predicción al contrastar los valores reales con los valores hipotéticos, procesados por la herramienta. Ciertamente, los resultados arrojados por el contraste de las muestras originadas en los registros de la base de datos determinan que los valores del atributo (TIEMPOEMPLO) correlacionados con el atributo (NIVEDUCMADRE), y contrastados con los resultados del algoritmo de la (PREDICION), tienen un éxito en el pronóstico efectivo del 74%, para la totalidad de los 4,656 registros valorados.

Estos hallazgos se asemejan a las conclusiones de otro estudio, realizado por el investigador Hoff (2003), quien supone que el mejor desarrollo de los adolescentes, en tareas de memoria y expresión, podría deberse a que los padres con una mayor educación crean ambientes intelectualmente más estimulantes para sus hijos. Es aquí donde valdría establecer si las entrevistas de trabajo, en donde la expresión oral es importante, están siendo motivo para perder un empleo. Agregando a lo anterior, el autor anota que los padres profesionales tienen una manera distinta de interactuar con sus hijos, especialmente en lo respecta con el lenguaje.

La correlación significativa del nivel de educación de la madre, va en la misma línea de los autores P. Marí-Klose, M. Marí-Klose, Granados, Gómez y Martínez (2009), quienes en un estudio que pretendía establecer la relación entre el nivel educativo de los padres y el acceso de sus hijos a la educación superior, establecieron que: "La educación de los padres es esencial en el logro escolar de los hijos" (p. 297); en la disertación se establece que el nivel de estudios de la madre es el factor que más influye en el devenir educativo de los hijos. Adicionalmente, sobre el éxito escolar y la deserción estudiantil los autores afirman:

El nivel educativo de la madre sí tiene un efecto consistente. Los jóvenes cuyas madres tienen un nivel de estudios universitarios tienen una razón de probabilidades 17 veces más alta de no haber abandonado los estudios de forma prematura que aquéllos cuya madre tiene estudios primarios. Los jóvenes cuya madre tiene estudios secundarios tienen una razón de probabilidades un 43% más alta. (p. 229).

El nivel educativo de los padres es la principal clave para los logros escolares que cosechen sus hijos. El perfil de los estudios de los progenitores, sobre todo el de las madres, es el factor más determinante en el éxito escolar de los hijos.

Adicionalmente, los resultados aquí discutidos, en relación con el nivel educativo de la madre, tienen concordancia con los autores Matute, Sanz, Guma, Rosseli y Ardila (2009), quienes establecen que las madres con educación superior emplean un vocabulario más rico y leen más a sus hijos, que las madres que sólo estudiaron la primaria; aseveran, además, que el nivel educativo de la madre afecta el desarrollo del lenguaje de los hijos e incide en el progreso de los procesos cognoscitivos, incluyendo la atención y la memoria.

En Colombia, según cifras del Dane a 2018, existen 12'768.157 jóvenes (entre 18 y 28 años), quienes representan el 27 por ciento de la población. De estos, 3'400.000 no tienen empleo, cifra que para expertos es una preocupante radiografía del desempleo juvenil en el país. (Portafolio, 2018, p. 1).

No obstante, los resultados predictivos de esta investigación abren un camino importante para establecer derroteros en toma de decisiones y retroalimentación de orden psicosocial, que permitan mejorar los indicadores predictivos futuros en pro del mejoramiento continuo.

| Conclusiones

Los resultados obtenidos en la presente investigación muestran que los recién egresados del programa de Ingeniería de Sistemas, de la Universidad de Caldas, entre los años 2012 y 2017 tienen una relación gradual, conforme al tiempo de consecución del primer empleo, y que este incremento es más acentuado para jóvenes, cuyo nivel educación de la madre tiene una escolaridad más alta. El perfil de los estudios de la madre progenitora, según las variables empleadas para esta disertación, se considera entonces como el factor más determinante en el éxito, referido al menor tiempo de búsqueda del primer empleo.

Decididamente, el nivel de alcance de los padres en la educación de sus hijos tiene efectos parentorios en las trayectorias profesionales de los recién egresados. El escenario presentado expone cómo el entorno y núcleo familiar, en donde la madre juega un papel esencial, es un determinante directo en los resultados educativos de la juventud de hoy. De ahí la importancia que adquieren las iniciativas gubernamentales, departamentales y universitarias orientadas a fortalecer el capital social, familiar y cultural, para producir resultados valederos a corto y mediano plazo.

De otro lado, el concepto y aprovechamiento de un modelo superior en la minería de datos, como lo es el paquete DBMS_PREDICTIVE_ANALYTICS, es visible en los resultados, ya que esta activa herramienta realiza la preparación automática de datos, mientras que el modelo está en construcción. Por consiguiente, añade más funcionalidad a la creación del simple modelo y da una experiencia superior fuera de confort del usuario. Este modelo no solo predice y lleva a cabo las tareas de minería de datos, sino que también prepara los datos para los algoritmos predictivos necesarios. Mejor es tomar medidas cuando se conoce el resultado probable, en lugar de simplemente esperar lo mejor y reportar los resultados después.

En esta investigación se realizó una contrastación usando técnicas automáticas de extracción de datos, buscando establecer los mejores predictores para el atributo tiempo primer empleo de un recién egresado, con el objeto de permitir focalizarse en los determinantes sociales familiares, y así descubrir que, en un entorno familiar, el nivel de educación de la madre es un estimulante directo que cataliza el esfuerzo laboral y académico de los hijos. El proceso de revelación de conocimiento tiene un análisis enmarcado en dos elementos básicos: el primero, descubrir con técnicas bayesianas los mejores predictores, y el segundo contrastar los datos reales con valores de predicción, otorgados por la herramienta. El estudio comprueba el valor objetivo original con el valor de predicción, en donde se concluye que los graduados al momento de conseguir su primer empleo tienen una relación directa con el nivel educativo de su progenitora.

Referencias

- Aponte F, Hoyos, J. y Monsalve, J. (septiembre-diciembre, 2012). Minería de usabilidad aplicada a plataformas virtuales de aprendizaje. *Revista Virtual Universidad Católica del Norte*, (37), 27-43. Recuperado de <http://revistavirtual.ucn.edu.co/index.php/RevistaUCN/article/view/386>
- Bedoya, O., Lopez, D. y Marulanda, C. (abril, 2018). Percepción de la adopción de la tecnología en los egresados de programas de humanidades de la Universidad de Caldas (Colombia) utilizando analítica predictiva en minería de datos. *Revista Espacios*, 39(10), 17. Recuperado de <http://www.revistaespacios.com/a18v39n10/18391017.html>
- Bedoya, O., López, M. y Marulanda, C. (septiembre-diciembre, 2016). Minería de datos en egresados de la Universidad de Caldas. *Revista Virtual Universidad Católica del Norte*, (49), 110-124. Recuperado de <http://revistavirtual.ucn.edu.co/index.php/RevistaUCN/article/view/800/1320>
- Buron, R. y Sausen, J. (abril, 2017). O papel da universidade na formação profissional na área da saúde. *Revista Espacios*, 38(30), 32. Recuperado de <https://www.revistaespacios.com/a17-v38n30/17383032.html>
- Chalaris, M., Gritzalis, S., Maragoudakis, M., Sgouropoulou, C. & Tsolakidis, A. (2014). Improving Quality of Educational Processes Providing New Knowledge Using Data Mining Techniques. *Procedia - Social And Behavioral Sciences*, 147, 390-397. doi: 10.1016/j.sbspro.2014.07.117.

- Gómez, A. y Suárez, C. (2012). *Sistemas de información* (3 ed.). México: Alfaomega.
- Han, J., Kamber, M. & Pei, J. (2012). *Data mining* (3 ed.). Waltham, United States: Morgan Kaufmann/Elsevier.
- Hand, D., Mannila, H. & Smyth, P. (2001). *Principles of data mining* (1 ed.). New Delhi, India: PHI Learning Private Limited.
- Hoff, E. (September-October, 2003). The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. *Child Development*, 74(5), 1368-1378.
- Marí-Klose, P., Marí-Klose, M., Granados, F., Gómez, C. y Martínez, A. (2009). *Informe de la Inclusión Social en España 2009*. Barcelona, España: Caixa Cataluña.
- Martelo, R., Acevedo, D. y Martelo, P. (2018). Análisis Multivariado aplicado a determinar factores clave de la deserción universitaria. *Revista Espacios*, 39(10), 13. Recuperado de <http://www.revistaespacios.com/a18v39n10/18391013.html>
- Marulanda, C., López, M. y Mejía, M. (febrero-mayo, 2017). Minería de datos en gestión del conocimiento de pymes de Colombia. *Revista Virtual Universidad Católica del Norte*, (50), 224-237. Recuperado de <http://revistavirtual.ucn.edu.co/index.php/RevistaUCN/article/view/821/1339>
- Matute, E., Sanz, A., Guma, E., Rosseli, M. y Ardila, A. (mayo, 2009). Influencia del nivel educativo de los padres, el tipo de escuela y el sexo en el desarrollo de la atención y la memoria. *Revista Latinoamericana de Psicología*, 41(2), 261-272.
- Medeiros, M. y Galvão, J. (2016). Educação e Rendimentos dos Ricos no Brasil. *Dados*, 59(2), 357-383. doi: 10.1590/00115258201680. Recuperado de <http://dx.doi.org/10.1590/00115258201680>
- Miranda, A. y Tirado, F. (octubre-diciembre, 2012). Las nuevas universidades: El fenómeno de comunidades de aprendizaje en línea. *Revista de la Educación Superior*, 41(164), 9-33. Recuperado de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-27602012000400001&lng=es&tlng=es
- Miranda, M. & Guzmán, J. (2017). Analysis of Dropouts of University Students using Data Mining Techniques. *Formación Universitaria*, 10(3), 61-68. Recuperado de <http://dx.doi.org/10.4067/S0718-50062017000300007>.
- Natek, S. & Zwilling, M. (October, 2014). Student data mining solution-knowledge management system related to higher education institutions. *Expert Systems with Applications* 41(14), 6400-6407.

- Portafolio, D. A. (16 de mayo de 2018). Desempleo juvenil sigue creciendo en Colombia. *Portafolio*, pp. A1.
- Rossetto, C. y Gonçalves, F. (2015). Equidade na Educação Superior no Brasil: Uma Análise Multinomial das Políticas Públicas de Acesso. *Dados*, 58(3), 791-824. Recuperado de <https://dx.doi.org/10.1590/00115258201559>
- Sánchez, C., Giraldo, L., Piedrahita, C., Bonet, I., Lochmuller, C., Tabares, M. y Peña, A. (marzo, 2018). Análisis comparativo entre: «el análisis exploratorio de datos» y los modelos de «árboles de decisión» y «kmeans» en el diagnóstico de la malignidad en algunos exámenes de cáncer de mama. Un estudio de caso. *Revista Espacios*, 39(28), 21. Recuperado de <http://www.revistaespacios.com/a18v39n28/18392821.html>
- Shaw, M., Subramaniam, C., Tan, G. & Welge, M. (mayo, 2001). Knowledge management and data mining for marketing. *Decision Support Systems*, 31(1), 127-137. doi: 10.1016/s0167-9236(00)00123-8
- Solarte, O. y Millán, M. (julio-diciembre, 2014). Propuesta para extender semánticamente el proceso de recuperación de información. *Revista EIA*, 11(22), 52-65.
- Timaran, R. (enero-junio, 2010). Una lectura sobre deserción universitaria en estudiantes de pregrado desde la perspectiva de la minería de datos. *Revista Científica Guillermo De Ockham*, 8(1). Recuperado de <http://www.redalyc.org/src/inicio/ArtPdfRed.jsp?iCve=105317327011>